

高等学校规划教材

关系数据库原理 习题及上机指导

刘方鑫 曲云尧 孟凡荣 编

XITI JI SHANGJIZHIDAO

中国矿业大学出版社

数据库原理习题及上机指导

10.311.13-818

出版社

高等学校规划教材

关系数据库原理习题及上机指导

刘方鑫 曲云尧 孟凡荣 编

中国矿业大学出版社

(苏)新登字第 010 号

内 容 简 介

本书是与《关系数据库原理及应用》(刘方鑫、曲云尧、孟凡荣编)一书配套的教学用书。全书概括地论述了关系数据库的基本概念、算法、程序设计方法和上机指导。结合教材内容每章备有习题及习题选解。全书共十二章,第一至第七章主要讨论了关系数据库原理及习题解;第八至第十一章讨论了dBASE III 的程序设计方法、技巧及习题解,第十二章给出上机指导。

本书根据教学大纲要求选择了一批具有典型性、综合性习题,并给出部分习题的解题方法与答案,为读者深入学习数据库技术提供有效的工具。

本书是计算机应用专业数据库原理及应用课程的教学用书,还可作为高等学校与计算机相关专业的教学用书,亦可供从事计算机、管理科学、数据处理的广大科技人员学习与参考。

责任编辑 胡玉雁

技术设计 关湘雯

高等学校规划教材

关系数据库原理习题及上机指导

刘方鑫 曲云尧 孟凡荣 编

中国矿业大学出版社出版

新华书店经销 中国矿业大学印刷厂印刷

开本 787×1092 毫米 1/16 印张 10.25 字数 248 千字

1993 年 12 月第一版 1993 年 12 月第一次印刷

印数 1—3000 册

ISBN 7-81040-179-3

TP·6

定价:4.60 元

前 言

关系数据库由于具有完备的理论基础、功能强、使用方便等优点,它已成为目前应用最广泛的数据库管理系统。目前出版的数据库方面教材一般只给出少量习题,且多属一般概念性问题,缺少解题方法指导等必要的内容。根据教学大纲要求我们编写了这本习题及上机指导。本书主要特点如下:

1. 本书概括地论述了关系数据库基本原理和应用实践。每章的内容提要给出了有关的基本概念和学习重点,为深入解题打下基础。

2. 选题围绕课程讲授重点、难点和开拓知识面的要求,选用多种类型。如选择题、问答题、计算题、分析判断题、综合设计题、填空题等,可满足不同读者的需要。本书部分习题给出题解和参考答案,以利于读者自学。

3. 注重理论联系实际。本书在论述关系数据库原理基础上,讨论了关系数据库语言dBASE III的程序设计及上机指导。

4. 本书既可与《关系数据库原理及应用》一书配套使用,又可独立作为教学辅助用书。读者在阅读内容提要基础上参照部分题解,可以进一步完成有一定难度的题解。这对深入掌握关系数据库原理及应用技术是十分必要的。

本书由刘方鑫任主编。刘方鑫编写第一、二、三、四、十、十一章,曲云尧编写第五、六、七章,孟凡荣编写第八、九及十二章。在编写过程中得到许多老师的关心和支持,编者在此谨向他们表示衷心地感谢。

由于我们水平有限和经验不足,因此书中疏漏谬误之处在所难免,敬请读者批评指正。

编者

1993.9

目 录

第一章 数据库系统引论	(1)
一、内容提要	(1)
二、习题	(1)
三、习题选解	(3)
第二章 数据库系统结构	(10)
一、内容提要.....	(10)
二、习题.....	(11)
三、习题选解.....	(12)
第三章 关系数据库方法	(18)
一、内容提要.....	(18)
二、习题.....	(20)
三、习题选解.....	(25)
第四章 关系规范化理论	(32)
一、内容提要.....	(32)
二、习题.....	(34)
三、习题选解.....	(36)
第五章 数据库保护	(44)
一、内容提要.....	(44)
二、习题.....	(46)
三、习题选解.....	(50)
第六章 数据库设计	(58)
一、内容提要.....	(58)
二、习题.....	(58)
三、习题选解.....	(61)
第七章 分布式数据库	(66)
一、内容提要.....	(66)
二、习题.....	(67)
三、习题选解.....	(69)
第八章 dBASE II 概述	(72)
一、内容提要.....	(72)
二、习题.....	(73)
三、习题选解.....	(74)

第九章 数据库的操作	(77)
一、内容提要	(77)
二、习题	(79)
三、习题选解	(90)
第十章 dBASE III 程序设计	(97)
一、内容提要	(97)
二、习题	(99)
三、习题选解	(100)
第十一章 几种常用微机关系数据库系统简介	(120)
一、内容提要	(120)
二、习题	(121)
三、习题选解	(121)
第十二章 上机指导	(125)
实验一 数据库文件的建立及函数的使用	(125)
实验二 数据库文件的查询、修改和拷贝	(131)
实验三 数据库文件的索引和多工作区的操作	(135)
实验四 统计和报表输出	(140)
实验五 程序设计	(144)
实验六 dBASE III 与高级语言之间的连接	(148)
实验七 综合练习	(154)
参考文献	(158)

第一章 数据库系统引论

一、内容提要

1. 数据库是存储在计算机系统内的一个通用化的、综合性的、有结构的数据集合；它具有最小的数据冗余度和较高的数据独立性、安全性和完整性；数据库的创建、运行和维护是在数据库管理系统控制下实现的。

2. 数据库系统的产生与发展。进入 60 年代末三个数据库系统(IMS 系统、DBTG 系统、关系数据库系统)相继出现，标志着数据管理技术进入数据库阶段。70 年代被称为数据库的年代，进入 80 年代数据库技术又有更大的发展。知识库、分布式数据库、面向对象数据库等技术成为人们关注的研究课题。

3. 数据库系统的特点：实现数据的集中化控制；数据的冗余度小；避免数据的不一致性；实现数据共享；提供数据保护；维护数据的独立性。

4. 实体集之间的联系有 1:1 联系、1:n 联系、m:n 联系。

5. 层次模型是按层次结构的形式来组织数据库数据的数据模型。它把实体及实体之间的联系表示为一棵有根的树，树的每个结点是一个记录类型，其基本联系方式是 1:n。层次结构的物理实现方法有邻接法、链接法等。

6. 网状模型指用网状数据结构表示实体和实体之间联系的数据模型。该模型允许一个子女记录型可以有多个双亲记录型，其基本联系是 m:n。网状结构的物理实现方法有邻接和链接混合法、链接法以及指针阵列法等。

7. 关系模型是采用二维表格形式的数据模型。即用二维表格表示实体及实体间的联系，表格是同类实体的各个属性的集合。任一实体模型均可用关系模型表示。关系模型具有坚实的理论基础、表达能力强、简单易学、数据独立性高等优点。

8. 数据库中常用的文件结构有串行文件、顺序文件、索引文件、B-树与 B⁺树、HASH 文件、倒排文件、表结构文件等。

二、习题

1. 什么是数据处理？什么是数据管理？数据管理的三个发展阶段是什么？

2. 什么是数据库？什么是数据库系统？什么是数据库管理系统？

3. 何谓数据的物理独立性与数据的逻辑独立性？

4. 试述与数据处理有关的三个世界(信息领域)之间的联系。

5. 试举出三个实例说明两个实体型之间具有 1:1, 1:n, m:n 的联系。

6. 层次模型有哪些特性？常用的物理实现方法有哪些？

7. 试述网状模型与层次模型的区别。

8. 举例说明复杂网状结构、简单环形结构、复杂环形结构的主要特征。

9. 什么是关系模型？关系模型的优点是什么？

10. 收集下列数据:

- 1) 5 个学生的学号、姓名、年龄、系别。
- 2) 5 门课的课程号、课程名、学时数、任课教师。

11. 设有关商品供应的信息如下:

厂商编号、厂商名称、所在城市、商品编号、商品名称、商品颜色、商品单价、库存量。试用关系模型分别表示厂商、商品、供应的有关信息。

12. 举例说明如何把网状模型转换为层次模型。

13. 设微机实体集包含微机编号、微机型号、微机名称等属性, 学生实体集包含学生编号、姓名、班级、年龄等属性。试用关系模型表示学生上机登记这种 $m:n$ 的联系。

14. 设一层次结构如图 1-1 所示; 其中包括的记录为:

A: A1、A2

B: B1、B2

C: C1、C3、C5、C7

D: D1、D4

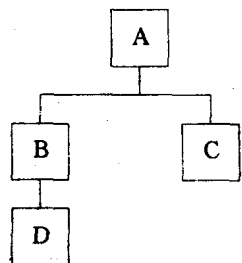


图1-1 层次模型实例

试用邻接法表示该层次模型的物理实现。

15. 在关系模型中如何表示实体集之间的联系?

16. 数据库中常用的文件结构有哪些?

17. 什么是顺序文件? 有哪些存储结构? 对分块查找和折半查找方法的查询效率做一比较。

18. 什么是索引顺序文件? 举例说明该文件的组织方法。

19. 什么是索引非顺序文件? 说明在该文件中进行查询、更新的方法。

20. B_树有哪些特点?

21. 一棵 $d=2$ 的 B_树实例如图 1-2 所示。

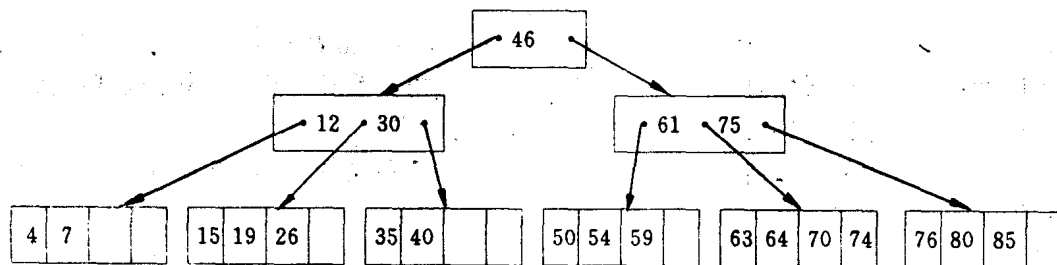


图1-2 $d=2$ 的B_树

1) 给出查找键值为 59 的节点的路径。

2) 在图 1-2 表示的 B_树中插入 37 和 72 后有何变化? (用图表示)

3) 画出在插入后的 B_树中删除 63 后的 B_树。

22. 什么是 B⁺树? B⁺树与 B_树的主要区别是什么?

23. 什么是 HASH 文件(亦称散列文件)? 举例说明几种散列算法(平方取中法、质数除余法、移位法)

24. 什么是倒排文件? 它有哪些类型?

25. 什么是表结构文件? 列举几种常用的表结构文件。

26. 定义并理解下列术语:

实体、实体集、属性、记录、字段、型、值、实体模型、数据模型、导航存取方法、入口点存取方法、稀疏索引、稠密索引。

27. 分析下列提法, 是否正确?

- 1) 一切层次结构和网状结构都可用关系模型表示。
- 2) 在数据库中如果完全消除了数据冗余, 则不会产生数据不一致性。
- 3) 在实际数据库系统中, 由于存在数据冗余, 则一定会产生数据不一致性。
- 4) 若数据库逻辑结构改变不需修改应用程序, 则称系统具有数据的物理独立性。

28. 分析下列提法, 选择正确的论述:

- 1) 一个网状结构可以变换为一个树形结构, 这种变换以存储空间为代价。
- 2) 顺序文件是按照数据到达文件的时间顺序依次连续地存储数据。
- 3) 评定一个文件性能的优劣主要看查询速度是否快。
- 4) 环(Ring)是把链文件的最末一个记录直接与头一个记录相连接的文件。

29. 分析下列提法, 指出错在哪里。

- 1) 从统计数据库(SDB)中可以取得个体数据。
- 2) 统计数据库的安全性完全依靠存取控制保护机构实现。
- 3) 在面向对象数据库中对象类型是若干不同对象的集合。
- 4) 属性值的域(Domain)仅包括数字数据。

30. 从供选择的答案中选出适当的字句, 完成下面关于数据库的叙述。

1) 从逻辑上看关系模型是用(a)表示记录类型, 用(b)表示记录类型之间的联系; 层次与网状模型是用(c)表示记录类型, 用(d)表示记录类型之间的联系。从物理上看关系是用(e), 层次与网状是用(f)实现两个文件之间的联系。

2) 目前已在计算机系统上建立的数据库管理系统: SYSTEM-R、DB2、INGRES、FOXBASE 是属于(g), IDMS 是属于(h), IMS 是属于(i)。

3) 数据库系统是由(j)、(k)、(l)、(m)、(n)组成; 而数据库应用系统是由(o)、(p)、(q)组成。

供选择的答案:

(1)表、(2)结点、(3)指针、(4)连线、(5)位置寻址、(6)相联寻址、(7)关系模型、(8)层次模型、(9)网状模型、(10)数据库系统、(11)数据库管理系统、(12)数据库、(13)数据库管理员、(14)计算机系统(软/硬件)、(15)批处理用户和终端用户、(16)应用程序系统。

三、习题选解

1. 数据处理是指对数据进行收集、组织、存储、加工、抽取和传播等一系列活动的总和。其目的是从大量的、原始数据中抽取、推导出对人们有价值的信息。

数据管理是指对数据的分类、组织、编码、存储、检索和维护等工作。数据管理是数据处理的中心环节。

数据管理技术随着计算机技术的发展而不断发展。数据管理技术可根据数据的独立性、数据的冗余度、数据间相互联系的方式、数据的安全性和完整性等特点划分为三个不同的发展阶段: 人工管理阶段(50年代中期以前)、文件系统阶段(50年代中至60年代后期)、数据

库系统阶段(60年代后期以来)。

3. 数据库中数据独立性包括数据的逻辑独立性和数据的物理独立性。

数据的逻辑独立性是指某一用户的逻辑数据文件(局部逻辑结构),因故发生变化而修改局部逻辑结构,或增加新的用户使数据库的全局逻辑结构改变时,其他用户的应用程序都无需修改。即数据库全局逻辑结构的改变,如数据定义的修改、增加新的数据类型、数据间联系的变更等,不影响应用程序的正常使用。因为系统提供了全局逻辑结构与局部逻辑结构间的映象功能。

数据的物理独立性是指数据的存储结构和存取方法改变时,数据的全局逻辑结构和应用程序都无需修改。因为在数据库中系统提供了存储结构与逻辑结构间的映象功能。

4. 数据库应能正确地反映现实世界中被管理事务的特征及其联系。如何将现实世界中的事务及其联系从计算机允许的形式反映到数据库中呢? 这里有个逐步转化过程,从现实世界→信息世界→计算机世界。通常称这三个与数据处理相关的领域为三个世界。

1) 现实世界是由各种各样实体组成,它是具体事务和抽象概念及其相互间联系的总和。实体由特性来表征和区分,特性中有些是该实体所固有的,有些则反映该实体与其它实体间的联系。具有主要特性相同的一类实体的集合称为实体集。实体集中的各个实体是通过实体标识符加以区别的。

2) 信息世界:现实世界中的实体通过人们的感官(眼、耳、鼻、舌、身)反映到人们的头脑中形成信息,组成信息世界。因此,信息世界是现实世界在人们头脑中的反映。在信息世界里用实体记录表示实体,用属性表示实体集的特性,用属性值表示一个实体的特性。

3) 计算机世界:人们对信息进行加工、整理、组织和抽象,然后用数字、文字或符号等有形的物理符号记录下来就形成了数据。因此,数据是信息世界中信息的数据化。但在计算机世界中一切信息只能用二进制数字表示,即在计算机世界中信息必须是数字化的。因此,计算机世界也成为数据世界。在计算机世界中用文件、记录、字段等术语表示现实世界的实体及其特性。现实世界实体及其间的联系,在数据世界中用数据模型表示。

5. 1) 两个实体型间有 1:1 联系见图 1-3a:

(1) 工厂与正厂长 (2) 学校与正校长 (3) 旅游者与护照



图 1-3a 1:1 联系

2) 两个实体型间有 1:n 联系见图 1-3b:

(1) 班级与学生 (2) 大学与教师 (3) 公司与职员



图 1-3b 1:n 联系

3)两个实体型间有 $m:n$ 联系见图 1-3c:

(1) 教师与学生 (2) 商店与商品 (3) 教练与运动员



图 1-3c $m:n$ 联系

7. 网状模型与层次模型的主要区别如下:

1)网状模型允许有一个以上结点无双亲,且至少有一个结点有多于一个的双亲,且必须为结点间的联系命名。

2)网状模型允许有复合链,即允许两个结点之间可以有二种或两种以上的联系。

3)网状模型对数据的搜索方式较为灵活,可以采用入口点存取方法及导航存取方法进行数据操作。

4)网状模型的基本联系是 $m:n$ 的,它可以表示各种类型的联系。

8. 用网状数据结构表示实体和实体间联系的模型称为网状模型。

下面结合实例说明复杂网状结构、简单环形结构和复杂环形结构的主要特征。

1)复杂网状结构具有 $m:n$ 的联系,图 1-4 给出具有 $m:n$ 联系的网状结构实例:

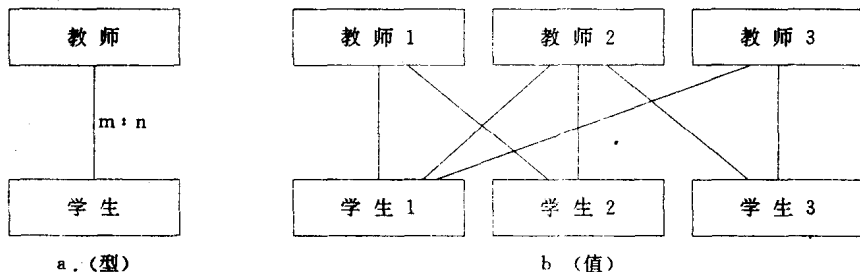


图1-4 复杂网状结构

该结构表示每个教师可与多个学生有教学联系,同样每个学生可与多个教师有听课联系。从型上看似乎与层次结构类似,其实不然。区分数据模型不能只从型上看,更主要的是从具体值的联系上看。从图 1-4b 可以看出:在结构中同一个结点的几个双亲结点都是同一种记录类型的值,且双亲与子女结点是 $m:n$ 联系。

2)简单环形结构是由一串首尾相接的记录型组成的环。记录类型间的联系是 $1:n$ 的。该环形结构又分为多个记录型的简单环和单个记录型的简单环,后者实例如图 1-5 所示。

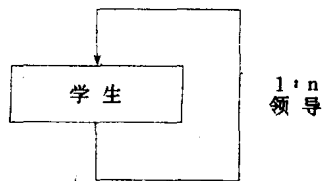


图1-5 单个记录型的简单环

对学生记录型,若要区别学生中领导与被领导的联系,则在学生会主席这个记录下有若干个从属的记录,它们之间有 $1:n$ 的联系(假设只有一个主席)。因此在单个记录型的简单环中只有一种记录型,即双亲记录和子女记录是同类的记录。

3)复杂环形结构是由一串首尾相接的记录型组成,记录型间的联系是 $m:n$ 。该结构也

有简单记录型复杂环和多记录型复杂环结构之分。后者的一个实例如图 1-6 所示。

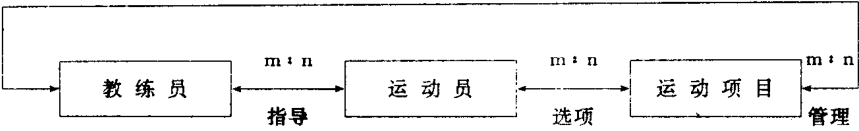


图 1-6 多记录型复杂环

在教练员记录型与运动员记录型间有“指导”联系。由于一个教练员可以指导多个运动员，一个运动员可以接受多个教练员指导（如基础训练、专项训练），因此，“指导”联系为 $m:n$ 。同样，在运动员与运动项目记录型间的“选项”联系也 $m:n$ 的。在教练员与运动项目记录型间有“管理”联系，即一个教练员（如体操教练）可以管理几个运动项目（如单杠、鞍马等），而一个运动项目也可以为多个教练员管理。因此，教练员与运动项目记录型间存在 $m:n$ 联系。

10. 关系模型是统一地用一种二维表格表示记录类型以及它们之间的联系。图 1-7 给出有关学生、课程及学生选课的关系实例。

1) 学生信息表见图 1-7a

学 号	姓 名	性 别	年 龄	系 别
920801	刘 一	男	20	计算机
920802	王 英	女	19	计算机
920901	张玉美	女	18	自动化
920910	李 四	男	19	自动化
920911	赵 五	男	22	自动化

图 1-7a 学生信息表

2) 课程信息表见图 1-7b

课 程 号	课 程 名	学 时 数	任课教师
C1001	微 机 原 理	60	AAA
C8008	关 系 数 据 库	70	BBB
A2002	控 制 原 理	70	CCC
C6006	C 语 言	40	DDD
E4001	英 语	100	EEE

图 1-7b 课程信息表

3) 学生选课表见图 1-7c

学 号	姓 名	课 程 号	课 程 名
920801	刘 一	C1001	微 机 原 理
920801	刘 一	C8008	关 系 数 据 库
920802	王 英	C6006	C 语 言
920802	王 英	E4001	英 语
920901	张玉美	A2002	控 制 原 理
920901	张玉美	E4001	英 语
920910	李 四	C1001	微 机 原 理
920910	李 四	E4001	英 语
920911	赵 五	A2002	控 制 原 理
920911	赵 五	C1001	微 机 原 理

图 1-7c 学生选课表

12. 网状模型可以通过分解等方法变换为层次模型。通常是将两个(或多个)双亲共有—个子女的结构变换为每个双亲各拥有一个相同的子女结构,如图1-8所示。

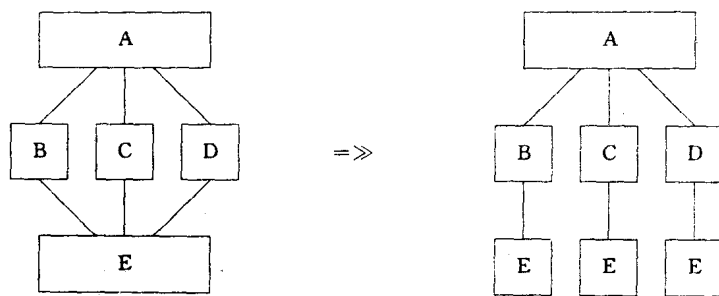


图1-8 网状模型变换为层次模型

显然,采用这种变换方式有更多的数据冗余。因此,由网状模型到层次模型的转换可借助于虚记录型的方法实现。一个虚记录实际上是一个指针,用它来指向真正的记录。在上面的例子中,变换后的层次模型中只保存一个结点E为记录型,而另外2个E结点是虚记录型。

14. 层次结构物理实现方法之一是邻接法。邻接法是用连续的物理顺序表示记录间的联系。即由根记录开始存放,按自顶向下、自左至右的顺序存储记录,按这种顺序排列的记录序列称为层次序列。图1-1层次模型的邻接存储结构如图1-9所示。

A1	B1	D1	D4	B2	D1	D4	C1	C3	C5	C7	A2	B1	D1	D4	B2	D1	D4	C1	C3	C5	C7
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

图1-9 层次模型的邻接存储结构

17. 顺序文件是将记录按关键字值的次序由小到大(升序)或由大到小(降序)顺序排列的文件。顺序文件的存储结构有:

- 1) 向量结构: 文件的逻辑结构与物理结构一致,记录按逻辑顺序连续存放。
- 2) 链结构: 一个文件以记录为单位分散在存储空间的不连续的位置上。每个记录都带有一个指向下一个记录的指针,最后一个记录带有链结束标记(Λ)。
- 3) 块链结构是向量结构和链结构的结合,在一个物理数据块中的记录按地址顺序连续存放,而块与块之间用指针链接以维持记录的逻辑顺序。

分块查找与折半查找是向量结构的顺序文件常用的两种查询方法。分块查找的效率与块的大小有关,计算表明最佳分块方法是每块记录个数 n 应为文件记录总数的平方根,即 $n = \sqrt{M}$ 。此时,查询一个记录的平均查找次数可表示为 $E_{\min} = \sqrt{M}$ 。而折半查找法查询一个记录的平均查找次数 $E = \log_2 M - 1$ 。当 M 较大时有 $(\log_2 M - 1) < \sqrt{M}$ 成立。所以,理论上讲当 M 较大时折半查找法的效率高于分块查找。

20. 索引是提高数据库存取效率的基本方法,多层索引常采用树型结构。1970年R. Bayer和E. McCreight提出一种外查找树结构称为B-树。这是一种动态调节的平衡树。它与普通树结构的区别在于对树中各节点的子女数目、每条路径的长度有一定的限制。B-树的主要特点:

- 1) 每个节点是一个物理块,每个物理块最多包含 $2d$ 个项(每个项是一个关键字)和 $2d + 1$ 个指针,其中 d 是一个与节点所含项数有关的因子称为阶。

- 2) 根结点至少具有一个项。除根结点外的每个结点至少包含 d 个项和 $d+1$ 个指针。
- 3) 具有 j 个项的结点, 有 $j+1$ 个子女。
- 4) 所有叶结点都在同一层上。通常叶结点不包含记录本身, 而是一些由指针和关键字组成的索引记录块。

21. 一棵 $d=2$ 的 B-树实例如图 1-2 所示。

1) 查找 $K_x=59$ 的节点路径:

从根节点开始比较 K_x 与节点中的关键字 $K_1, K_2, \dots, K_j$ 。由于 $K_x > \text{根节点 } K_1(46)$, 则沿着 P_j (右指针) 所指向的子树 (右子树) 继续查找。由于 $K_x < 61$, 则沿着左指针所指向的子树 (叶节点) 继续查找, 该节点中查到与 $K_x=59$ 相同的项, 则查询结束。若按上述路径直到叶节点仍无 K_x 值, 则表示要查找的记录不在该 B-树中。

2) 对图 1-2 所示的 $d=2$ 的 B-树插入键值为 37 和 72 的项后如图 1-10 所示。

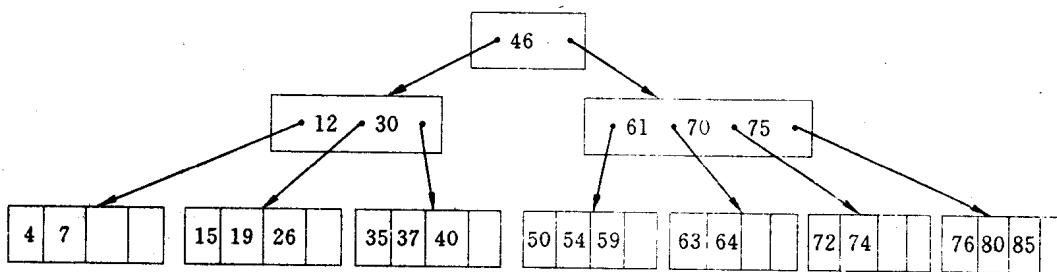


图1-10 插入操作后的 $d=2$ 的B-树

3) 在图 1-10 所示 B-树上, 删除 $K_x=63$ 的项后如图 1-11 所示。

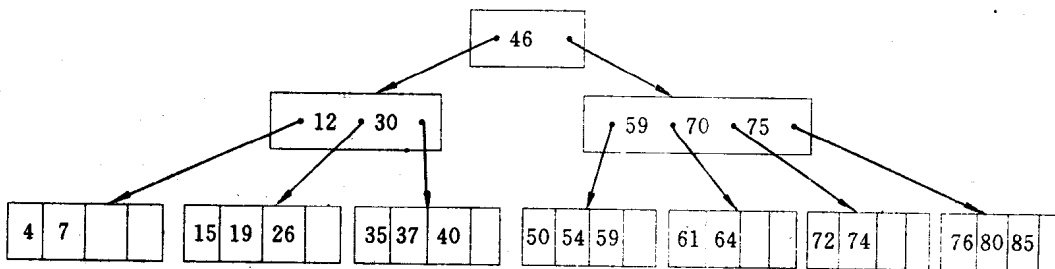


图1-11 插入操作后的 $d=2$ 的B-树

23. HASH 文件是利用散列法进行组织的文件。散列法是通过一种散列函数或算法把记录的关键字转换成记录地址的方法。HASH 文件也称散列文件、随机文件或直接文件。下面举例说明几种常用的散列算法。

1) 平方取中法: 先把数字化的键值平方, 取平方值的中间 n 位数通过变换把它调整到存储桶号范围内作为地址。

例如, 设桶的个数是 4000, 某一记录的键值 $K=244447$ 。则 $K^2=597655809$, 取出从左边数第 4 位到第 7 位数字得 6558, 再按 4000 范围进行调整 $\lfloor (6558 * 4000) / 10 \rfloor = 2000$, 这就是对应的存储桶号。

2) 质数除余法: 取最接近桶数的质数作为除数去除数字化的键值, 其余数作为桶号。若

余数的变化范围超过存储桶数目时,则按比例把它调整到存储桶数目的范围内。

例如,设桶数为 6010,取质数为 6007,要转换的键值是 $K=274238$,那么 $274238/6007$ 的余数为 3923,就是对应的桶号。

3) 移位相加法:将数字化的键值分成左右两段(也可以参照桶数目分成多段),使之相向移动对齐并相应于地址长度,然后相加并对结果进行调整成为散列地址。

例如,设桶数为 2000,要转换的键值为 43633872,分成左右两段相加:取后三位数 235 作为桶号。

$$\begin{array}{r} 4363 \\ + 3872 \\ \hline 8235 \end{array}$$

24. 倒排文件是借助于多重键来组织文件的。在倒排文件中用主键以外的属性作为辅助键组织辅助索引表(称为倒排表)。倒排表由辅助键和指针组成,指针可以是记录的地址,也可以是主关键字。建立辅助索引的主要目的是适应根据次关键字的查询,通过它可以根据一个或多个数据项进行查询,以提高查询速度。

倒排文件根据文件“倒排”的程度不同可分为三类:

- 1) 完全倒排文件:文件中对应于每个数据项都有一个索引,即每个数据项都是键。
- 2) 部分倒排文件:文件中只有用部分数据项作键,建立相应的索引表。
- 3) 辅助索引文件:辅助索引文件同正规文件一样仍保留主键,由主键的顺序决定记录的存放顺序,在主索引表中给出记录地址。此外,它还建立了辅助索引,在辅助索引中给出主键值。

28. 正确的论述是(a),其余几种提法不准确。如(b)所论述的文件应该是串型文件,而顺序文件是根据键值经过排序整理过的串行文件。(c)评价一个文件性能的优劣主要看对文件进行存取、增删操作是否灵活方便,查询速度是否快,存储空间利用率是否高等。(d)环是把链文件的最末一个记录通过指针与头一个记录相连的文件,而不是直接相连。

30. 在实用数据库系统中经常使用三类数据模型:层次、网状、关系。层次模型是树形结构,网状模型是丛结构,它们在逻辑上用结点表示记录类型,用连线表示记录类型之间的联系,而在物理上用指针来实现两个文件之间的联系。关系模型采用二维表,既表示记录类型,又表示记录类型间的联系。而在物理上则用相联寻址实现两个表格之间的联系。

本题的答案是:

- (a) : (1), (b) : (1), (c) : (2), (d) : (3), (e) : (6), (f) : (3),
(g) : (7), (h) : (9), (i) : (8), (j) : (11), (k) : (12), (l) : (13)
(m) : (14), (n) : (15), (o) : (10), (p) : (15), (q) : (16)。

第二章 数据库系统结构

一、内容提要

1. 数据库的三级模式结构。根据美国 ANSI/X3/SPARC 数据库管理系统研究组 1975 年发表的报告,一个数据库结构从逻辑上可以划分为三级模式:外部模式、概念模式、内部模式。这个报告是一个有关数据库系统结构模型标准化的重要文件,三级模式结构被许多数据库系统采用。

1) 外部模式是用户观念下的数据结构的逻辑描述,表示了用户所理解的实体、实体属性和实体间的联系。它是用户与数据库间的接口。

2) 概念模式是数据库全局逻辑结构的描述。

3) 内部模式是对数据库中数据的物理结构的描述。

外部模式和概念模式之间通过外部模式/概念模式映象进行变换;概念模式和内部模式之间通过概念模式/内部模式映象进行变换。这些模式间的映象为数据库中数据的独立性提供有效的支持。

2. 数据库的抽象层次。数据库系统的三级模式结构定义了数据库的三个抽象层次:物理数据库、概念数据库和逻辑数据库。

1) 物理数据库是以内部模式为框架的数据库,是物理存储设备上实际存储着的数据集合。

2) 概念数据库是以概念模式为框架的数据库,是数据库的整体逻辑表示。

3) 逻辑数据库是以外部模式为框架的数据库,是用户所看到和使用的数据库。

3. 数据库系统管理员 DBA。

DBA 是数据资源管理机构的一组人员(一般不多于 2 人),负责全面地管理和控制数据库系统。主要职责有:

1) 参与数据库系统的设计与建立;决定数据库的信息内容和结构;决定数据库的存储结构和存取策略;定义数据的安全性要求和完整性约束条件。

2) 监督和控制数据库的使用和运行。

3) 定义和实施适当的后援和恢复策略,负责数据库的恢复工作。

4) 改进数据库的性能,对数据库进行重组和重构。

4. 数据库管理系统 DBMS 是一个负责数据库的定义、建立、操作、管理和维护的软件系统。它建立在操作系统的基础上,对数据库进行统一的管理和控制。

5. DBMS 的主要功能有:

1) 数据库定义功能;2) 数据库管理功能;3) 数据库的建立与维护;4) 故障恢复功能;5) 通讯功能。

6. DBMS 的组成。

1) 系统运行控制程序组: 系统总控程序、存取控制程序、并发控制程序、保密控制程序、数据完整性控制程序、数据库存取访问和更新程序、通讯控制程序。

2) 语言处理程序组: 数据库各级模式的语言处理程序、数据操纵语言 DML 处理程序、终端查询语言解释程序、数据库控制语言解释程序。

3) 数据库日常管理程序: 数据装入程序、数据恢复程序、数据库重构程序、统计分析程序、工作日志程序、转储程序。

7. 数据描述语言 DDL 用于描述数据库的各级模式。根据所定义的不同级别的模式, 数据描述语言分为模式 DDL、子模式 DDL、存储模式 DDL。

1) 模式 DDL 用于对数据库全局逻辑结构的具体描述的数据描述语言。

2) 子模式 DDL 是供用户描述所用部分数据的局部逻辑结构的数据描述语言。

3) 存储模式 DDL 是用于对数据库存储模式进行描述的数据描述语言。

8. 数据操作语言 DML 供用户对数据库进行查询、更新操作的语言。

常用的数据操纵语言有两种基本类型: 主语言型数据操纵语言和自含语言型数据操纵语言。

9. 关系数据库系统 System R 的分层结构, 包括视图、基表、存储文件三个层次。

10. System R 的体系结构。该系统的数据库管理系统由两个子系统组成: 一个是关系数据库系统 RDS, 简称数据子系统; 另一个是研究存储系统 RSS, 简称存储子系统。

二、习题

1. 何谓数据库的三级模式结构, 这种结构的优点是什么?

2. 什么是外部模式、概念模式、内部模式? 它们之间有何联系?

3. 什么是逻辑数据库、概念数据库和物理数据库? 它们之间有何联系?

4. DBA 的主要职责是什么?

5. DBMS 的主要功能是什么?

6. DBMS 有哪些程序组组成?

7. 什么是数据描述语言, 有哪几种?

8. 如何用关系数据描述语言定义关系数据库模式? 结合下面给出的三个关系实例用数据描述语言给出关系数据库模式的描述。这三个关系实例是:

学生关系 $S(SNO, SNAME, SDEPARTMENT, SAGE)$, $KEY = SNO$;

课程关系 $C(CNO, CNAME, PCNO)$, $KEY = CNO$, 其中 PCNO 是某课程的先导课程号, 并假定每门课只有一门先导课。

学生选课关系 $SC(SNO, CNO, GRADE)$, $KEY = (SNO, CNO)$ 。

9. 存储描述语言的主要功能是什么?

10. 数据操纵语言的功能有哪些?

11. 什么是内容定址法、数据关联法?

12. 什么是主语言型数据操纵语言? 如何把 DML 嵌入宿主语言?

13. 何谓自含语言型数据操纵语言? 有何特点?

14. 试述关系数据语言的特点。

15. 试画图说明用户通过应用程序在数据库中修改一个记录的主要过程。

16. 什么是数据字典? 说明数据字典的内容。